Testing Multidimensional Models of Youth Civic Engagement:

Model Comparisons, Measurement Invariance, and Age Differences

Laura Wray-Lake

University of Rochester

Aaron Metzger

West Virginia University

Amy K. Syvertsen

Search Institute

Citation:
Wray-Lake, L., Metzger, A., & Syvertsen, A. K. (2017). Testing multidimensional models of youth civic engagement: Model comparisons, measurement invariance, and age differences. *Applied Developmental Science, 21.4*, 266-284. doi:0.1080/10888691.2016.1205495

This version does not include changes made in the final proofing stages of publication. Link to published article:
http://www.tandfonline.com/doi/full/10.1080/10888691.2016.1205495

Acknowledgements: This work was funded by a grant from the John Templeton Foundation. The opinions expressed in this presentation are those of the authors and do not necessarily reflect the values of the John Templeton Foundation. We are grateful to the schools and the youth who participated in the study. We thank Jennifer Shubert, Ben Oosterhoff, Brian Lake, Lauren Alvis, Rebecca Olson, Michael Warren, Chen-Yu Wu, and Maura Shramko for their assistance in this investigation, and Todd Little for his comments.

Abstract

Despite recognition that youth civic engagement is multidimensional, different modeling approaches are rarely compared or tested for measurement invariance. Using a diverse sample of 2,467 elementary, middle, and high school-aged youth, we measured eight dimensions of civic engagement: social responsibility values, informal helping, political beliefs, civic skills, environmental behavior, volunteering, voting intentions, and news consumption. We compared correlated unidimensional factors, higher-order factor, and bifactor models and tested for measurement invariance and latent mean differences by age. The correlated unidimensional factors model best fit the data, yet higher-order and bifactor models fit adequately. Metric and scalar invariance was found across models. Latent means varied depending on the dimension of civic engagement and the multidimensional model examined. Findings favor the correlated unidimensional factors model; implications of each model are discussed. This study informs future research on youth civic engagement and has broad relevance for any developmental scientist studying a multidimensional construct.

KEYWORDS: positive youth development; measure development; civic development; structural equation modeling; measurement model comparisons; multiple group models; age differences

Testing Multidimensional Models of Youth Civic Engagement:

Model Comparisons, Measurement Invariance, and Age Differences

Developmental science seeks to understand complex, multidimensional phenomena (Overton, 2015). Multidimensionality implies that there are numerous parts to a larger whole and is indicated when two or more separable scales fit under the same conceptual umbrella. Many developmental constructs are recognized as multidimensional, and multidimensional thinking has ushered in theoretical and empirical advancements in areas such as prosociality (Padilla-Walker & Carlo, 2014), personality (Chen, Hayes, Carver, Laurenceau, & Zhang, 2012), temperament (Rothbart, 2012), empathy (Decety, 2012), well-being (Ryff, 2014), and positive youth development (Bowers, Li, Kiely, Brittian, Lerner, & Lerner, 2010). However, there are numerous ways to model multidimensionality of a given construct, and each approach comes with different theoretical assumptions and empirical implications. Too often, measurement models are not explicitly tied to theory or empirically justified by comparing alternative models.

In this paper, we illustrate how to assess multidimensionality using youth civic engagement as an example. This study has three aims: (a) Describe and empirically compare multidimensional models to determine which model(s) best capture youth civic engagement, (b) Examine measurement invariance by age to evaluate the quality of measures, and (c) Test mean differences by age to provide evidence of age differences in civic engagement across late childhood and adolescence. We examine three distinct approaches to modeling multidimensionality – correlated unidimensional factors, higher-order factor, and bifactor models – and compare these to a unidimensional single latent variable model. Findings inform future theory and research on youth civic engagement and illustrate broader measurement and conceptual implications of different modeling approaches.

A Multidimensional View of Civic Engagement

Civic engagement is defined as the behaviors, values, knowledge, and skills that comprise political and prosocial contributions to community and society (Sherrod & Lauckhardt, 2009). Youth express commitments to society in vastly different ways; thus, taking a multidimensional view of civic engagement is necessary for fully understanding youth's experiences. Multidimensional conceptualizations of civic engagement are being articulated with increasing clarity (Amnå, 2012; Haste & Hogan, 2006; Sherrod & Lauckhardt, 2009) and advance the field by recognizing that civic behavior is multifaceted and includes actions as diverse as voting, volunteering, activism, and environmental conservation. It is conceptually important to measure multiple civic behaviors because youth gravitate towards different civic actions based on background, contexts, interests, and opportunities (Wray-Lake & Sloper, 2015). Contemporary scholarship has also recognized that sociocognitive components (e.g., values, skills, knowledge) are essential to defining and understanding civic engagement (Flanagan, 2013; Metzger & Smetana, 2009; Voight & Torney-Purta, 2013). In the current study, we embrace the view that civic engagement has multiple dimensions. Whereas there are probably more ways to express civic engagement than any single study can measure, we examine eight dimensions of youth civic engagement that include various behaviors and sociocognitive components and are thought to be relatively common and accessible to a wide age range of youth: social responsibility values, informal helping, political beliefs, civic skills, environmental behavior, volunteering, voting intentions, and news consumption.

Multidimensional measurement approaches can inform research on youth civic engagement in three key ways: First, multidimensional measurement models can inform fundamental definitional and conceptual questions about civic engagement. Despite some consensus that civic engagement is multidimensional, there is debate about whether dimensions of civic engagement form a coherent whole (e.g., Zaff et al., 2010) or whether civic engagement is comprised of distinct separable components (e.g., Geller, Voight, Wegman, & Nation, 2013). As summarized in Table 1 and described below, this question has implications for how civic engagement is conceptualized in theoretical models and how it is promoted in policy and practice. Distinct measurement models have not been empirically tested simultaneously or compared with rigor, and the current study addresses this research gap.

Second, multidimensional measurement models may provide more nuanced information about age differences in civic engagement and thus contribute to emerging developmental theory of youth civic engagement (Lerner, Wang, Champine, Warren, & Erickson, 2014). Empirical research on age differences in youth civic engagement is limited to single dimensions such as informal helping (Carlo, Crockett, Randall, & Roesch, 2007) or social responsibility values (Wray-Lake, Syvertsen, & Flanagan, 2016), but lacks cohesive examination across multiple ages and measures. Our findings contribute to developmental research by examining age differences in civic engagement levels across late childhood through adolescence. As described further below, different measurement models may offer divergent conclusions about age differences.

Third, civic measurement is an active area of study, marked by several notable attempts to disseminate comprehensive survey measures, yet more measurement work is needed, particularly from a developmental perspective (Flanagan, Syvertsen, & Stout, 2007; Syvertsen, Wray-Lake, & Metzger, 2015; Zaff et al., 2010; cf. Torney-Purta, Cabrera, Roohr, Liu, & Rios, 2015). We help fill this void by testing convergent validity of civic measures and examining measurement invariance of youth civic engagement across ages. To provide evidence of convergent validity, we link civic engagement measures to purpose. Purpose is a long-term commitment to a goal that is larger than the self, and through civic engagement youth often find a sense of purpose (Damon, Menon, & Bronk, 2003; Malin, Ballard, & Damon, 2015). Civic engagement is broadly conceptualized in this work; thus a general civic engagement factor may be strongly linked to purpose. Specific dimensions of civic engagement may differentially predict purpose and provide evidence of divergent validity, a possibility that remains untested.

Distinct Approaches to Modeling Multidimensional Constructs

We articulate three distinct approaches to multidimensional measurement models in the context of studying youth civic engagement. Although there are other viable multidimensional measurement model approaches – such as multiple higher-order factors, multitrait-multimethod models, or mixture models – we examine three models that are theoretically justified for civic engagement and can be empirically compared: a correlated unidimensional factor model, higher-order factor model, and bifactor model. Previous studies have compared these three models to demonstrate how empirical models align with distinct theoretical notions (Brunner, Nagy, & Wilhelm, 2012; Reise, 2012): These tutorials walk scholars through the mathematical details behind each model. Our paper highlights the conceptual rationale underlying each model and discusses the implications of each model for understanding youth civic engagement.

Approach 1: Correlated Unidimensional Factors

One approach to thinking about a multidimensional construct entails assuming that multiple specific dimensions of a construct fit together conceptually but are best measured distinctly. That is, the multidimensional construct is a conceptual idea but not a measurable construct. This idea is statistically represented as a correlated unidimensional factor model, also termed a first-order factor model, in which a construct is conceived of as distinct, but related, pieces. Regarding civic engagement, the construct would be modeled by first-order latent factors such as social responsibility values, informal helping, political beliefs, civic skills, environmental behaviors, volunteering, voting intentions, and news consumption (see Figure 2). An assumption of the correlated unidimensional factor model is that individual latent variables are conceptually related. Importantly, however, the model does not require that the direction or magnitude of these correlations be specified in advance (Brunner et al., 2012), allowing for flexibility in how much and in what ways the dimensions are related. For scholars who philosophically value specificity, correlated unidimensional factors have a strong theoretical appeal. Correlated unidimensional factor models assume that differences between dimensions are more important to study than their shared variance.

Many civic scholars embrace the correlational unidimensional factor approach by operationalizing civic engagement as multiple separate variables (e.g., Crocetti, Jahromi, & Meeus, 2012; Geller et al., 2013; Metzger & Smetana, 2009; Mahatmya & Lohman, 2012). Research has found that youth engage in different types of civic activities based on their motivation, gender, family and community bonds, identity status, and sociocognitive judgments (Metzger & Smetana, 2009; Wray-Lake & Sloper, 2015). This work showcases how developmental theory can be advanced by documenting specificity of processes for different civic dimensions and offering evidence of discriminant validity in predicting key outcomes. Research has also highlighted consistency of correlates across dimensions of civic engagement (e.g., Boyd et al., 2011), providing evidence for processes that generalize across dimensions. In our study, separate estimation of eight dimensions of civic engagement allows for identifying differences and similarities in age differences across these dimensions. Pinpointing specificity can move theory, research, and applied efforts beyond generalities and potentially address the critique that positive developmental research often relies on the balance principle and concludes that "good leads to good" (Heider, 1958). Correlated unidimensional factors models may be able to offer precise applied recommendations about specific pathways to specific constructs.

Approach 2: Higher-Order Factor

A second approach to multidimensionality is to model a higher-order construct, which represents the theoretical notion that a construct is multifaceted and hierarchically structured. Higher-order factor models build directly on a correlated unidimensional factor model by starting with unique first-order latent variables and defining a second-order factor as the shared variance among first-order constructs (see Figure 3). In other words, the higher-order factor explains intercorrelations among the first-order constructs (Chen, Sousa, & West, 2005; Chen et al., 2012). If a researcher's conceptualization of a construct truly lies at the global level, then examining separate individual factors (as in a correlated unidimensional factors model) would be unsatisfying because specific dimensions are considered only parts of the whole. Likewise, by emphasizing shared variance among dimensions, distinctions are relegated to the background. The higher-order factor model views a multidimensional construct as an integrated whole, and the theoretical benefits to this approach lie in providing broad understanding of a construct.

Modeling civic engagement as a higher-order factor model implies that civic engagement is best represented by the shared variance across multiple behavioral and sociocognitive components. The theoretical assumption is that civic engagement is best defined as an individual's coordination of beliefs, actions, and skills into a measurable whole, and examining one or a few dimensions of civic engagement would lead to partial understanding. Zaff and colleagues (2010) discussed the construct of *Active Engaged Citizenship* as a developmental integration of civic actions, skills, connections, and duties, arguing that fully engaged individuals will possess elevated levels of *all* components. Some scholars have adopted this view implicitly by recognizing distinct dimensions of civic engagement but then summing or averaging them (e.g., Kahne & Sporte, 2008; Lenzi et al., 2012). Civic engagement conceived of as a higher-order factor model could enhance parsimony by avoiding redundancy in predicted pathways across multiple dimensions of civic engagement that share variance. By identifying factors that explain variance in civic engagement as a whole, a higher-order model could lead to more generalizable conclusions about developmental processes and avoid piecemeal analyses that emphasize distinctions and may capitalize on elevated Type-I error rates that result from multiple statistical tests. The higher-order approach may have practical utility for policy, because it is more parsimonious to consider civic engagement as an integrated whole.

Approach 3: Bifactor

There are merits to viewing a multidimensional construct in terms of its specific and general parts. Typically these ideas are competing, yet a bifactor modeling approach allows researchers to simultaneously examine both. Bifactor modeling partitions variance into a general latent variable that accounts for commonality among items (e.g., general civic engagement factor) and a set of specific latent variables (e.g., volunteering, political behavior, social responsibility values) comprised of unique variance over and above the general factor (Chen et al., 2012). The underlying assumptions are that civic engagement is a measurable construct understood by integrating across dimensions, and each dimension can also be uniquely understood. A bifactor model is typically estimated by loading all items onto a general factor and all the items onto their respective specific factors (see Figure 4). The general and specific latent factors are assumed to be orthogonal (Reise, 2012). Although some argue that bifactor models differ more conceptually than mathematically from second-order models (Little, 2013), the main mathematical difference pertains to the proportionality constraint: bifactor models allow for more variability in ratios between general and specific factor loadings for an item, whereas loadings are implicitly constrained to be proportional in a higher-order model (Gignac, 2016). When items load differently onto the general versus the specific factor, a bifactor model will typically fit the data better than a higher-order factor model (Gignac, 2016). In addition, bifactor models are better able to determine whether specific factors exist beyond the general factor and offer a more straightforward way to examine unique predictors or outcomes of specific factors (Chen et al., 2012). Like the correlated unidimensional factors model, bifactor models can offer evidence of discriminant validity. By allowing for prediction of differences between dimensions *and* predicting shared variance, bifactor models can contribute to theory in acknowledging both specific and general processes and can offer a refreshing balance of general and specific recommendations for policy and practice.

Measurement Invariance by Age

Measurement invariance tests (i.e., equivalence of parameters such as factor loadings, intercepts, error variances) across groups such as age, gender, ethnicity, or socioeconomic status are important for determining whether measures are equally valid across groups and represent a crucial first step to achieve before means or structural paths can be compared (Kline, 2015). Although thorough measurement studies should test measurement invariance across multiple groups of interest, invariance by age is absolutely essential from a developmental perspective: Assumptions of measurement variance by age must be met before drawing inferences about cross-sectional age differences or longitudinal age-related change (Horn & McArdle, 1992; Widaman, Ferrer, & Conger, 2010).

Different levels of invariance can be reached (Meredith, 1993). Metric invariance refers to equivalence of factor loadings across groups or time; this is also called weak invariance and is seen as a minimal requirement for comparing constructs across groups or time (Little, 2013). Scalar invariance refers to equality of intercepts, and having both metric and scalar invariance is considered strong invariance. Having equivalent item-level means across groups after factoring out shared variance due to factor loadings is a prerequisite to interpreting latent variable mean differences (Little, 2013). Intercept differences could indicate problems with measurement or reflect substantive nuance in specific indicators not being captured by the construct. For example, if a global construct fails to account for the meaningful variance in a specific dimension of civic engagement, unique variance could get pushed down to indicator level. Finally, in higher-order and bifactor models, invariance of the first-order disturbances (i.e., variances of specific or lower-order factors) must be tested (Chen et al., 2005). This invariance gives confidence that lower-order latent variables (in a higher-order model) or specific latent variables (in a bifactor model) are equivalent across groups or time.

Measurement invariance tests by age have not been featured in published work for youth civic engagement measures, with the exception of Zaff and colleagues' (2010) longitudinal measurement invariance tests. Without broad measurement of civic engagement across ages and tests of measurement invariance by age, developmental research on youth civic engagement will stall, as this is an essential first step to answering fundamental questions about developmental change and processes. Cross-sectional studies can contribute to this work by testing if the measurement model structure is reasonably equivalent across age, and then testing for latent mean differences by age.

Age Differences in Levels of Civic Engagement

Research on age differences in civic engagement is sparse, and our large and diverse (albeit cross-sectional) sample of children and adolescents can add new evidence to this area.

The normative growth hypothesis posits that civic engagement increases across adolescence in concert with normative age-related growth in identity, autonomy, and exposure to contextual opportunities (Wray-Lake, Rote, Victorino, & Benavides, 2014). A mid-adolescence decline may typify certain civic constructs, such as social responsibility values, which show declines during middle adolescence and higher levels in elementary and high school (Wray-Lake et al., 2016). Related constructs of prosocial behavior and social trust have shown declines across adolescence (Carlo et al., 2007; Flanagan & Stout, 2010). Given divergent findings and a lack of research, further investigation of age patterns in youth civic engagement is sorely needed.

Importantly, however, each multidimensional approach to modeling civic engagement may offer a different conclusion for developmental theory. For example, the correlated unidimensional factor model should provide the most nuanced age differences, as specific dimensions may follow different age patterns, yet broad conclusions regarding an overall age-related pattern could not be discerned in this model. The higher-order factor model should provide evidence for age differences at the broadest level, yet distinctions may be washed out or identified through intercept invariance tests. In a bifactor model, general and specific conclusions can be drawn, yet age findings may differ most dramatically from other models, given that variance is partitioned into shared and unique components.

The Current Study

Table 1 provides a summary of the different theoretical implications and practical value of each multidimensional approach. To advance the study of multidimensional constructs in developmental research in general and for civic engagement specifically, we address three aims. Aim 1empircally compares different measurement models to determine which model best represents our data on youth civic engagement. Three distinct multidimensional approaches are compared to each other and to a unidimensional model (Model 0, see Table 1 and Figure 1). In Model 0, civic engagement is assumed to be a unidimensional construct, with all indicators equally loading onto the general factor. We hypothesize that youth civic engagement is multidimensional, but make no predictions about which multidimensional model best fits the data. Aim 2 is to evaluate the quality of civic engagement measures for developmental research. We test metric and scalar measurement invariance by age (elementary, middle school, and high school ages) for all empirically justified models, and convergent validity is tested by linking civic engagement to purpose. Aim 3 documents mean age differences in civic engagement. We interpret latent mean differences for all empirically justified models, anticipating that each model may provide different conclusions for developmental theory on civic engagement.

METHOD

Youth ages 8 to 20 (M = 13.4, SD = 2.7) enrolled in grades 4-12 were recruited from 17 schools in three socioeconomically, racially, and ethnically diverse regions of the United States: metropolitan California, urban Minnesota, and rural West Virginia. Student eligibility for free and reduced lunch ranged from 26 to 95% (M = 60%) across schools indicating substantial economic diversity across schools (National Center for Education Statistics, 2014).

Youth (N = 2,475) completed paper and pencil surveys in the classroom. Eight cases were excluded due to problematic response patterns; they were identified as multivariate outliers and failed attention check questions. Thus, 2,467 youth were used in analyses (56% female). The sample was 51% White, 30% Hispanic or Latino/a, 10% Black or African American, 7% Asian, 4% American Indian or Alaska Native, 2% Native Hawaiian or Other Pacific Islander, and 9% identified as another race-ethnicity. Youth reported primary caregivers' level of education on a 3-point scale: high school or below (Mothers: 27%; Fathers: 31%; Other Adult: 21%), some college (Mothers: 15%; Fathers: 14%; Other Adult: 14%), and college graduate or higher (Mothers: 35%; Fathers: 33%; Other Adult: 20%). Some youth reported they "Don't Know" their parenting adults' education level: Mothers: 23%; Fathers: 23%; Other Adult: 45%. Regarding financial strain, 10% of youth reported their family had a hard time buying the things they need, 34% reported their family as having just enough money for the things they need, 48% reported their family had no problem buying the things they need, and 9% reported their family had enough money to buy almost anything they wanted. In addition, 8% of youth reported being first-generation immigrants and 32% reported being second-generation immigrants.

To reduce participant burden, we employed a three-form planned missing design (Graham, 2012). See Online Appendix A for more detail. Planned missing data is controlled by the researcher and thus missing completely at random (MCAR). Survey versions were equally distributed across age, gender, ethnicity, parent education, and site (all chi-square tests were not significant). The Principal Components Method was used to incorporate principal components as auxiliary variables in the FIML missing data model (Howard, Rhemtulla, & Little, 2015).

Measures

Civic measures were drawn from existing sources and heavily adapted (Flanagan et al., 2007; Kahne, Middaugh, & Schutjer-Mance, 2005) or newly written but conceptually based on extant work (see Syvertsen et al., 2015). Items were examined in an iterative process involving multi-phase interviews, cognitive interviews, and a pilot survey study of 213 elementary, middle, and high school youth to ensure the developmental appropriateness of measures of youth across ages. Items were identically worded across age groups, with careful attention for easy readability and interpretation by the youngest participants. Omega coefficients are reported for elementary (E), middle (M), and high school (H) ages in the measure descriptions below.

Social responsibility values were measured with 4 items: "It is important to me to:": "consider the needs of other people," "help those who are less fortunate," "make sure that all people are treated fairly," "think about how my actions affect people in the future." Responses ranged from *Not at all important* (1) to *Extremely important* (5), ω = .63 (E), .79 (M), .77 (H).

Informal helping was measured with 6 items assessing the frequency of everyday forms of helping, including: standing up for a classmate that was being picked on; helping a classmate with homework; doing household chores such as cleaning, cooking, or yard work; sharing school supplies with peers; helping a neighbor with projects for no pay; and, babysitting for no pay. Responses ranged from *Never* (1) to *Very often* (5), ω = .52 (E), .65 (M), .64 (H).

Political beliefs were measured with 2 items assessing beliefs about actions: "People should keep up with current events and politics," "People should take part in a protest or rally to help change a law that they disagree with." Response options were: *Doesn't matter* (1), *Maybe should* (2), *Probably should* (3), *Mostly should* (4), *Definitely should* (5), r = .34 (E), .45 (M), and .49 (H), all ps < .001.

Youth self-rated their ability to perform six civic skills: "create a plan to address a problem," "get other people to care about a problem," "express my views to others in-person or in writing," "contact someone in a leadership position about a problem," "listen to conflicting viewpoints and identify where they agree and disagree," and, "summarize what another person said to make sure I understood." Response options were: *I definitely can't* (1), *Probably can't* (2), *Unsure if I can* (3), *Probably can* (4), and *Definitely can* (5), ω = .72 (E), .84 (M), .84 (H).

Three items gauged environmental behaviors: "I turn off electronics when I'm not using them," "I try to limit how much paper I use," and "I conserve water by taking shorter showers." Response options ranged from *Never* (1) to *Very often* (5), ω = .66 (E), .69 (M), .69 (H).

Volunteering was measured with a single item: "In a typical MONTH, about how many hours do you spend volunteering (not part of a class project, graduation requirement, or court-ordered requirement) to help other people or to help make your community a better place?" Response options ranged from *0 hours* (0) to *5+ hours* (6).

Voting intentions were measured by asking: "Have you ever done or plan to do the following? Vote in national elections." Response options were: *I wouldn't do this* (1), *Probably wouldn't do this* (2), *Unsure* (3), *Probably will do this* (4), *Will do or have already done this* (5).

News consumption was measured by a single item measuring how often participants "access information about politics and current events on TV, the radio, in the newspaper, or on news websites" in a typical week. Response options ranged from *Never* (1) to *Very often* (5).

Three items measured purpose, or commitment to a future goal that is larger than the self, adapted from Benson and Scales (2009): "I believe I am going to make a difference in the world," "I feel a sense of purpose in life," and "I have plans for my future." Response options ranged from *Not at all like me* (1) to *Very much like me* (5), ω = .66 (E), .69 (M), .74 (H).

For multigroup age models, the sample was divided into elementary (4-5th graders, n = 512), middle (6-8th graders, n = 813), and high school-aged youth (9-12th graders, n = 1,135). Grade and age were highly correlated (r = .81, p < .001). Although we are most interested in age developmentally, we chose school level as a proxy for age group because this offered a straightforward grouping that was preferable to selecting an arbitrary age cut-offs.

Analytic Plan

A series of structural equation models tested the factor structure of civic engagement. To account for the nested nature of individuals in schools, all models included school as a cluster variable (n = 17). Due to analysis of clustered data, analyses employed maximum likelihood estimation with robust standard errors (i.e., MLR). Separate confirmatory factor analyses (CFAs) were run to model dimensions of civic engagement as: a unidimensional model (Model 0), correlated unidimensional factors (Model 1), a higher-order factor model (Model 2), and as a bifactor model (Model 3). Scaling was done using the effects coding method, so that latent means could be reported (Little, Slegers, & Card, 2006; see Table 2). The two-item political beliefs construct was additionally scaled by constraining factor loadings to be equal. Standard model fit criteria were used, including chi-square tests, root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), and the Comparative Fit Index (CFI). Acceptable model fit values are .05 or lower for RMSEA and SRMR and .90 or higher for CFI, with .95 and higher preferred (Kline, 2015). Models were statistically compared based on chi-square difference tests using the Satorra-Bentler scaled method (required when using MLR estimation) to evaluate their relative fit (Satorra & Bentler, 1999). To establish validity of civic measures and compare models, we next included purpose as a dependent variable in each model.

For each model that showed acceptable model fit, multiple group analysis was utilized to assess measurement invariance across three age groups (elementary, middle school, high school). Across multiple group models, we scaled latent variables using the fixed factor method (i.e., latent variable variances fixed to 1) to facilitate comparison of factor loadings and means (Little, 2013). Elementary was the reference group. In conducting these tests, we primarily relied on CFI change to evaluate invariance, as chi-square difference tests have been found to be too liberal in assessment of invariance for large samples. Aligned with current recommendations, a CFI difference of .01 or greater was interpreted as substantively important and has also been considered a measure of effect size for invariance tests (Cheung & Rensvold, 2002; Little, 2013). To test metric invariance (i.e., equivalence of factor loadings), we compared a configural model where factor loadings freely varied across groups to a model with factor loadings constrained to be equal. Using the fixed factor scaling method, latent variable variances were fixed in Elementary and freely estimated in Middle and High groups. To test scalar invariance (i.e., equivalence of intercepts), we compared the metric invariance model from the previous step (in which factor loadings were fixed and intercepts were freed) to a model where intercepts were constrained to be equal across groups. In this step, latent variable means were fixed at 0 in Elementary and freely estimated in Middle and High Groups. In the higher-order factor model, tests of metric and scalar invariance were conducted separately at the first-order and higher-order levels (Chen et al., 2005). In the higher-order and bifactor models, we additionally tested for invariance of first-order factor disturbances (Chen et al., 2005). When the CFI change was greater than .01, models were considered non-invariant and modification indices were used to identify sources of differences and parameters were freed until partial invariance was reached. Satorra-Bentler scaled chi-square difference tests were used to evaluate individually freed parameters using a Bonferroni correction of $p = .002$ (Little, 2013). In the final step, we interpreted latent mean differences. The latent mean of the referent group (Elementary) was fixed to zero and latent means for Middle and High groups represented deviations from the Elementary mean. Latent mean differences were evaluated by examining significance of parameter estimates and significant chi square difference if parameters were constrained (using $p < .002$).

Results for each model are presented in the following sequence: (a) CFA for the full sample, (b) links from the civic engagement model to purpose, (c) tests of measurement invariance, and (d) interpretation of latent means by age.

RESULTS

Bivariate correlations among all study variables are presented in Table 2. Associations among variables were stronger between items from the same construct than across constructs.

Model 0: Unidimensional Model

The unidimensional model was estimated by loading all 24 items onto a single latent variable. Based on multiple fit indices, this model was a poor fit to the data, *MLR* $\chi^2(252) = 3659.01$, $p < .001$, CFI = .69, RMSEA = .074 (90% CI: .072 - .076), SRMR = .080 and fit significantly worse than the multidimensional models (Table 3). Standardized factor loadings ranged from .23 to .72, with five loadings below .4 (Table 4). Thus, results indicated that a single unidimensional construct of civic engagement was not viable in our data. Given the poor fit, no further analyses were conducted with this model.

Model 1: Correlated Unidimensional Factors

A first step to modeling correlated unidimensional factors was verifying that each dimension of civic engagement was distinct. In an exploratory factor analysis estimating 1 to 13 factors, results showed that an 8-factor model was better fitting than models with fewer factors and models with 9 or more factors did not offer meaningful improvements.

A CFA estimated 5 latent variables (social responsibility values, informal helping, political beliefs, civic skills, and environmental behaviors) and 3 manifest single-item variables (volunteering, voting intentions, and news consumption). The model provided a good fit to the data, *MLR* $\chi^2(228) = 831.29$, $p < .001$, CFI = .95, RMSEA = .033 (90% CI: .030 - .035), SRMR = .033 (Table 3). Standardized factor loadings ranging from .43 to .76 (*p*s < .001; Table 4). Covariances among dimensions of civic engagement were positive and significant, ranging from .09 to .59. The smallest correlations were with environmental behaviors and volunteering (r = .10), voting intentions (r = .09), and news consumption (r = .15). The largest correlations were between social responsibility values and informal helping (r = .59) and between informal helping and civic skills (r = .57). The majority of other correlations ranged from .2 to .4 (Table 5).

Next, we linked the unidimensional factors to purpose to examine validity. As expected, social responsibility values, informal helping, political beliefs, civic skills, and voting intentions were positively associated with youth purpose. However, environmental behavior, volunteering, and news consumption were not associated with purpose (Table 6).

To test metric invariance by age, the configural model with all parameters free to vary across groups was compared to a model with factor loading constrained to be equal (Table 7). Based on a ΔCFI of .002, we concluded that factor loadings did not significantly differ by age. Likewise, the test for scalar invariance was non-significant, ΔCFI = .005.

We next examined mean differences by age in the 8 civic engagement constructs (see Table 8). Elementary youth were lower on informal helping compared to middle and high school youth; the latter groups did not differ. High school youth were higher on political beliefs and civic skills compared to elementary and middle school youth; the latter groups did not differ. An unexpected pattern for environmental behavior showed that elementary youth were highest, followed by middle and then high school youth (all three groups differed). There were no age differences in social responsibility values, volunteering, voting, or news consumption.

Model 2: Higher-Order Factor Model

A second-order CFA included the 5 latent and 3 manifest civic variables from Model 1 as indicators of a second-order civic engagement factor. Model fit was acceptable, $MLR\chi^2(248)$ = 1012.24 $p < .001$, CFI = .93, RMSEA = .035 (90% CI: .033 - .038), SRMR = .039, although this model fit the data worse than Model 1, indicating the correlated unidimensional factors model fit better than the second-order model (Table 3). First-order loadings ranged from .43 to .76, and second-order factor loadings ranged from .33 to .76 (all ps < .001; Table 4).

The higher-order civic engagement latent variable was strongly positively associated with youth purpose (Table 5).

Metric invariance was tested separately for lower- and higher-order factor loadings. Based on ΔCFIs of .005, and .001, respectively, we concluded factor loadings were equivalent across groups. Regarding scalar invariance at the item level, intercepts were determined to be invariant based on ΔCFI of .003. In examining the intercepts of the first-order latent variables, a significant CFI change (ΔCFI=.018) revealed non-invariance. Based on modification indices, 4 first-order latent factor intercepts were freed across groups (Table 7). The environmental behavior intercept was freed for middle and high school youth, indicating again that elementary youth were highest on environmental behavior, followed by middle and then high school youth (Table 8). Social responsibility values were freed for high school, indicating that these values were lower for high school youth compared to elementary and middle school youth. Informal helping was freed for middle and high school youth, whose means were constrained to be equal: Elementary youth reported lower informal helping than middle and high school youth.

For the higher-order model, invariance of disturbances of first-order factors was tested. Disturbances were determined invariant, based on a non-significant chi square and ΔCFI of 0.

The higher-order civic engagement factor was highest in high school compared to elementary and middle school-aged youth; the latter two groups did not differ (Table 8).

Model 3: Bifactor Model

In Model 3, we estimated a bifactor model that included the 8 specific civic engagement constructs (5 latent, 3 manifest as described above) and a general civic engagement factor comprised of loadings from the 24 items. The model fit the data acceptably well, *MLR* $\chi^2(232) = 870.35$, $p < .001$, CFI = .942, RMSEA = .033 (90% CI: .031 - .036), SRMR = .036. Standardized factor loadings for the general civic engagement factor ranged from .18 to .58, with 8 loadings below .40 (*p*s < .001, Table 4). Bifactor models assess both shared and unique variance, which typically leads to specific factor loadings being lower than in Models 1 and 2 (Chen et al., 2012). The bifactor model showed fit better than the higher-order model, but fit worse than the correlated unidimensional factors model (Table 3).

The general civic engagement factor was again strongly positively associated with purpose. Social responsibility values, informal helping, civic skills, and voting intentions were positively associated with purpose, but political beliefs, environmental behavior, volunteering, and news consumption were not (Table 6).

The metric invariance test showed that factor loadings were equivalent across age groups, ΔCFI = .007. The test of scalar variance showed that intercepts were equivalent, ΔCFI = .006. First-order factor disturbances were also equivalent across groups, ΔCFI = .001 (Table 7).

Similar to the higher-order model, high school-aged youth were higher on general civic engagement than middle and elementary school-aged youth (Table 8). No age differences emerged for informal helping, political beliefs, or civic skills. As in other models, environmental behaviors were lower at successive ages. Social responsibility values were lowest in high school, followed by middle school, and then elementary school (all three groups differed).

DISCUSSION

Our results demonstrate that youth civic engagement is a multidimensional construct, supporting contemporary conceptual thinking (e.g., Sherrod & Lauckhardt, 2009). In comparing three distinct multidimensional models, the correlated unidimensional factors model best fit our data. However, notably, higher-order and bifactor models also provided good fit to the data. Evidence for metric and scalar measurement invariance by age and convergent validity (although it varied by model) provided evidence that our measures are well-positioned for further examining developmental questions related to civic engagement. Our study offers new evidence of cross-sectional age differences in civic engagement, but divergent findings across models also illustrate that model selection should be guided by both theory and empirical tests.

Evaluating Models

Comparing three multidimensional models of civic engagement to a unidimensional single latent variable provided strong evidence for the multidimensional structure of civic engagement. Scholars from across disciplines have described the multidimensional nature of civic engagement among youth and adults (Amnå, 2012; Haste & Hogan, 2006; Sherrod & Lauckhardt, 2009), and our work provides empirical support for these conceptual claims. In comparing the other three multidimensional models, the best-fitting model was the correlated unidimensional factors model, with the bifactor model the second-best fitting and the higher-order factor model fitting least well of the multidimensional models. The implications of these model comparisons extend to developmental science broadly and to research on civic engagement specifically.

Based on model fit criteria, each model fit the data acceptably well and thus could be empirically justified, which is common across studies of multidimensional constructs (Brunner et al., 2012; Reise, 2012). Our results align with similar model comparisons of internalizing symptoms by Reise (2012), who concluded that empirical differences were small between models, and thus in practice, model choice should be guided by theory and study goals. Thus, an important take-away point for scholars interested in multidimensional constructs is to prioritize conceptual rationale in model selection, carefully considering study purpose and goals for theory and practice along with empirical advantages and disadvantages (see Table 1). Moreover, given the relevance of model comparisons for clarifying definitions and conceptualizations of a construct, multiple models should be explicitly empirically compared (Wang et al., 2015). For example, our results favor the correlated unidimensional factors model and conflict with the model presented by Zaff and colleagues (2010), who found evidence for a higher-order factor model of civic engagement using a smaller set of civic constructs and data from the 4-H longitudinal study of mostly White, mostly middle class youth ages 14 to 16. Notably, if we had not empirically compared models, our findings could have aligned with Zaff et al.'s (2010) results based on a reasonably good fitting higher-order model. Thus, we underscore the need to replicate measurement model findings by conducting rigorous measurement work in other samples in order to best inform theory and practice.

Regarding implications of model comparisons for research on civic engagement, the best-fitting correlated unidimensional factors model implies that civic engagement is a conceptual idea measured by individual constructs that capture distinct dimensions of civic engagement. The magnitude of correlations among the individual latent variables varied substantially, explaining why the higher-order and bifactor models were not adequately capturing these covariances with a latent civic factor. Substantively, the correlated unidimensional factors model is well-poised to spur greater specificity in understanding civic engagement and related developmental processes. Recent work is already advancing our understanding of specificity in correlates of youth civic engagement (Crochetti et al.; Duke et al., 2009; Metzger & Ferris, 2013; Wray-Lake & Sloper, 2015). Greater use of the correlated unidimensional factors approach will advance civic engagement theory that articulates different precursors and outcomes of distinct dimensions of civic engagement (Metzger & Smetana, 2009). In turn, nuanced recommendations will likely emerge for practitioners in applied settings and civic education curricula. Practitioners tend to crave specificity in recommendations; the utility of the correlated unidimensional factors model lies in illuminating specific levers of change for each dimension of civic engagement.

Given other viable models, researchers should consider the limitations of the correlated unidimensional factors approach alongside its strengths. The large number of estimated parameters means greater model complexity and including covariates could easily lead to testing a dizzying number of model parameters, causing problems with model convergence and increasing probability of Type-I error. Complex and unexpected results could emerge from models with multiple dependent variables, heightening the potential for researchers to generate post-hoc, atheoretical explanations for findings. In areas such as youth civic engagement where theorizing about specific pathways is lacking, complicated and contradictory results could lead to more confusion than clarity about developmental processes.

The higher-order factor model revealed age differences in the civic engagement factor that align with the normative growth hypothesis that civic engagement increases with age; this idea has received only partial support in previous work (Wray-Lake et al., 2015) and has not been examined to a great extent. It would be difficult to ascertain this sort of holistic understanding of civic engagement with the correlated unidimensional factors model. Often, research starts with broad, general understanding and then proceeds to specific nuances (Metzger, Oosterhoff, Palmer, & Ferris, 2014). For a complex construct such as civic engagement, results related to a higher-order factor provide a straightforward and simpler way to discuss a phenomenon that policymakers and the public may more easily digest. This higher-order conception of civic engagement may be too general for practitioners, however, as it can be hard to know how to promote such as a diffuse construct. A reason to be cautious about higher-order approaches is that they can mask important variations occurring among specific dimensions of a construct. If theory suggests that specific dimensions of a construct should show different age patterns, aggregating across dimensions could wash out unique variance and preclude understanding of specific developmental changes. In this study, we would likely not use a higher-order factor model due to its poor relative fit compared to the other multidimensional models. However, particularly given Zaff et al.'s (2010) previous work, it is worth considering this model in further work on civic engagement and for other multidimensional constructs.

A bifactor model has not been examined with youth civic engagement to our knowledge, and in our study, the bifactor model offered a better fit than the higher-order factor model. Unlike a higher-order factor model, a bifactor model can be useful in cases, like in our study, where each dimension contributes differently to the general factor (Chen et al., 2012). The bifactor model offers an ideal circumstance in conferring advantages of both general and specific approaches, increasing broad understanding of civic engagement and illuminating specificity.

Despite their conceptual and empirical appeal, bifactor models are uncommon in the literature. Perhaps bifactor models are difficult to estimate, given high model complexity and possible instability of model structure over time. In bifactor models, alternative explanations could easily apply to the general factor, which may represent unwelcomed sources of shared variance such as social desirability or positivity bias. A consistent and expected pattern of results gave us confidence in the interpretation of our general factor as civic engagement. For example, the pattern of associations linking specific and general factors to purpose replicated across correlated unidimensional and higher-order models. Also, the same age pattern emerged for the civic engagement factor in the bifactor and higher-order models. More broadly, interpretations of specific dimensions in a bifactor model must acknowledge that these dimensions represent unique variance only. Thus, results with specific dimensions in bifactor models may not always align with models that do not remove shared variance. Our divergent age findings across models for social responsibility values, described further below, are a good example of this. Our bifactor model included several low loadings, which can be substantively interesting in illustrating the amount of variance an item contributes to a general versus specific factor (Gignac, 2016). However, when a latent variable such as informal helping has several low factor loadings, this pattern may suggest lower reliability of the specific construct. In summary, bifactor models should be further explored for use with youth civic engagement and other constructs. We would feel confident using the bifactor model in our data, particularly in situations where the research questions and study goals necessitate use of specific and general civic engagement constructs.

Evaluating Measures

After accounting for clustering by school, assumptions of metric and scalar invariance were largely met across models. Measurement invariance tests are essential to advancing developmental research. In our case, finding invariance gave us confidence to proceed with examining age differences in levels of constructs and suggests a path forward for future work to examine developmental change longitudinally using these measures. Our work contributes to youth civic engagement research by identifying reliable, valid, and developmentally appropriate civic engagement measures for youth from late childhood through late adolescence.

However, our measures are not a panacea, and limitations should be noted. Political activism is an important aspect of civic engagement (Kirshner, 2015), but these kinds of political behaviors are expected to be relatively rare, particularly among younger children, and due to concerns about survey length and item complexity, we asked political activism measures of middle and high school youth but not elementary. At a more basic level, each specific dimension of a construct has to load strongly and positively onto a broad construct before a higher-order approach can be viable. Dimensions of civic engagement that we did not examine – such as political activism – may negatively correlate with other dimensions and invalidate a higher-order model. Reliability coefficients were lower on most dimensions for elementary youth. This pattern is not surprising, as developmental research has long shown that measurement reliability for multiple constructs increases with age, and differences are attributed to cognitive development such as growth in abstract thinking, memory, and language (e.g., Edelbrock, Costello, Dulcan, Kalas, & Conover, 1985). We do not view this pattern as problematic in our data, given that reliabilities for younger youth were approaching acceptable standards of .70 and measurement invariance tests suggested equivalence across age groups. However, there may be room to further improve the measurement of these constructs for our youngest participants.

Finally, with respect to convergent validity tests, not all dimensions of civic engagement were positive associated with purpose. Purpose may only be related to certain dimensions of civic engagement such as social responsibility, informal helping, civic skills, and voting intentions. Civic engagement as a general factor was consistently linked to purpose, and likewise, scholarship conceptually linking civic engagement to purpose conceives of civic engagement broadly (Damon et al., 2003; Malin et al., 2015). Overall, validity tests may need to become more refined for civic engagement as theory and research get more specific and if the correlated unidimensional model is used. Currently, few papers on civic measurement conduct validity tests and much more work is needed in this area (Torney-Purta et al., 2015).

Evaluating Age Differences

The correlated unidimensional factors model, identified as best-fitting, also showed the greatest specificity in age differences for dimensions of civic engagement. Higher-order and bifactor models showed that their broad civic engagement factor was higher in high school compared to younger ages. Thus, results from these two models provide some support for the normative growth hypothesis that civic engagement increases across adolescence (Wray-Lake et al., 2015). Since the shared variance for civic engagement was captured in a general civic engagement factor for these two models, age differences in specific dimensions mostly reflect deviations from this general pattern. On the other hand, the correlated unidimensional factors model provides the most direct test of which civic dimensions follow this pattern: Informal helping, political beliefs, and civic skills are higher for high school youth in this model. Not surprisingly, these findings largely disappear in the other two models because this shared variance is moved to the general/higher-order civic engagement factor (with the exception of civic skills that remain higher). Thus, age differences across models mostly tell a similar story.

In some cases, age findings were inconsistent across models, most notably for social responsibility values. Previous research has found declines in social responsibility values across adolescence (Wray-Lake et al., 2016). Our results show that social responsibility values follow this pattern of being lower at older ages only when shared variance with other dimensions of civic engagement are factored out (i.e., in the higher-order and bifactor models). In the correlated unidimensional factors model, social responsibility values were highly correlated with informal helping and civic skills. It is likely that factoring out this shared variance leaves meaningful unique variance for social responsibility; alternatively, factoring out shared variance could substantially change the construct. Clearly, removing shared variance changes the association between age and social responsibility values. Given the notable differences in interpretation and alignment with previous work, the social responsibility findings illustrate the importance of using conceptual rationale to guide selection of multidimensional models to avoid cherry-picking findings. On the other hand, consistent results across models for the other civic constructs we measured in this study are encouraging and increase confidence in interpretation of findings. For instance, the consistent finding that environmental behaviors were lower in each successive age group is informative in suggesting that older adolescents may be less inclined toward everyday environmental conservation. This finding merits replication in other studies and with other measures, given that this pattern was unexpected based on the normative growth hypothesis. This pattern could be due to our measure's narrow focus on conservation.

Limitations and Future Directions

Strengths of our study include the large socioeconomically, racially, ethnic, and geographically diverse sample that spanned a wide age range. Our study is primarily limited by its cross-sectional design. Cross-sectional age comparisons offer a useful starting point, yet longitudinal data are necessary for examining developmental change. Three broad age groups are likely insufficiently nuanced to capture complex (e.g., non-linear) age patterns. Model selection criteria certainly influence conclusions drawn. We evaluated invariance tests using a change in CFI rule, which is conservative in reducing Type-I error (Cheung & Rensvold, 2002). More differences would have emerged on factor loadings and on intercepts had we used different criteria, such as the chi-square difference tests or a more sensitive CFI rule (Meade, Johnson, & Braddy, 2008). Furthermore, Gignac (2016) argues that fit indices that add a penalty for model complexity can better determine differences between higher-order and bifactor models. More uniformity in applying model evaluation criteria in the field would better facilitate ability to draw conclusions across studies. Other models may offer different ways to think about multidimensionality, such as mixture models that recognize individual variability in response patterns and can determine that facets work together in different ways for different people. This kind of idiographic approach to multidimensionality is important to consider and can offer a nuanced look at civic engagement (e.g., Voight & Torney-Purta, 2013). Finally, it was beyond the scope of this paper to test for measurement invariance across other key subgroups (e.g., gender, ethnicity), but future work should conduct these important tests to better understand the viability of these measures for assessing civic engagement across groups.

In a complex world with complex constructs, it is critical for researchers to consider the implications of different approaches for modeling multidimensional phenomena. Our illustration demonstrates the utility of empirical tests for model comparisons but also shows that model selection should be driven by theoretical rationale. As a field, we should better recognize that the findings of any single study should not constitute the basis for theory or practical solutions, given that findings are sensitive to small nuances in model selection. Replication and accumulation of evidence across studies and across methodological approaches is important for drawing firm conclusions. Our paper offers a set of valid, reliable, and age invariant measures of civic engagement that, particularly when modeled multidimensionally, can enhance understanding of development in this domain. In studying multidimensional constructs, we urge scholars to take comfort in the now classic statistical adage, “All models are wrong, but some are useful” (Box & Draper, 1976, p. 424). There is room for both general and specific approaches to multidimensional constructs, and different approaches advance theory in distinct ways and produce different kinds of evidence of value for policy and practice.

References

Amnå, E. (2012). How is civic engagement developed over time? Emerging answers from a multidisciplinary field. *Journal of Adolescence, 35,* 611-627. doi:10.2478/s13374-012-0024-1.

Benson, P. L. & Scales, P. C. (2009). The definition and preliminary measurement of thriving in adolescence. *The Journal of Positive Psychology*, *4*(1), 85-104, doi: 10.1080/17439760802399240

Bowers, E. P., Li, Y., Kiely, M. K., Brittian, A., Lerner, J. V., & Lerner, R. M. (2010). The five Cs model of positive youth development: A longitudinal analysis of confirmatory factor structure and measurement invariance. *Journal of Youth and Adolescence*, *39*(7), 720-735. doi:10.1007/s10964-010-9530-9

Box, G. E. P., & Draper, N. R. (1987). *Empirical model-building and response surfaces*. Hoboken, NJ: Wiley.

Boyd, M. J., Zaff, J. F., Phelps, E., Weiner, M. B., & Lerner, R. M. (2011). The relationship between adolescents' news media use and civic engagement: The indirect effect of interpersonal communication with parents. *Journal of Adolescence*, *34*, 1167-1179. doi:10.1016/j.adolescence.2011.07.004

Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality, 80*(4), 796-846. doi: 10.1111/j.1467-6494.2011.00749.x

Carlo, G., Crockett, L. J., Randall, B. A., & Roesch, S. C. (2007). A latent growth curve analysis of prosocial behavior among rural adolescents. *Journal of Research on Adolescence, 17*(2), 301–324. doi:10.1111/j.1532-7795.2007.00524.x

Chen, F. F., Hayes, A., Carver, C. S., Laurenceau, J. P., & Zhang, Z. (2012). Modeling general and specific variance in multifaceted constructs: A comparison of the bifactor model to other approaches. *Journal of Personality*, *80*(1), 219-251. doi:10.1111/j.1467-6494.2011.00739.x

Chen, F. F., Sousa, K. H., & West, S. G. (2005). Teacher's corner: Testing measurement invariance of second-order factor models. *Structural Equation Modeling*, *12*(3), 471-492. doi: 10.1207/s15328007sem1203_7

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness of fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9*, 23 –255. doi: 10.1207/S15328007SEM0902_5

Crocetti, E., Jahromi, P., & Meeus, W. (2012). Identity and civic engagement in adolescence. *Journal of Adolescence*, *35*(3), 521-532. doi: 10.1016/j.adolescence.2011.08.003

Damon, W. (2009). *The path to purpose: How young people find their calling in life*. Simon and Schuster.

Damon, W., Menon, J., & Bronk, K. (2003). The development of purpose during adolescence. *Applied Developmental Science*, *7*(3), 119-128.

Decety, J. (2012). *Empathy: From bench to bedside*. Cambridge, MA: MIT Press.

Duke, N. N., Skay, C. L., Pettingell, S. L., & Borowsky, I. W. (2009). From adolescent connections to social capital: Predictors of civic engagement in young adulthood. *Journal of Adolescent Health*, *44*(2), 161-168. doi: 10.1016/j.jadohealth.2008.07.007

Edelbrock, C., Costello, A. J., Dulcan, M. K., Kalas, R., & Conover, N. C. (1985). Age differences in the reliability of the psychiatric interview of the child. *Child development*, *56,* 265-275. doi:1. Retrieved from http://www.jstor.org/stable/1130193 doi:1

Flanagan, C. A. (2013). *Teenage citizens: The political theories of the young*. Harvard University

Press.

Flanagan, C. A., & Stout, M. (2010). Developmental patterns of social trust between early and late adolescence: Age and school climate effects. *Journal of Research on Adolescence, 20*(3), 748-773. doi: 10.1111/j.1532-7795.2010.00658.x

Flanagan, C. A., Syvertsen, A. K., & Stout, M. D. (2007). *Civic measurement models: Tapping adolescents' civic engagement*. CIRCLE Working Paper 55. College Park, MD: Center for Information and Research on Civic Learning and Engagement (CIRCLE).

Geller, J. D., Voight, A., Wegman, H., & Nation, M. (2013). How do varying types of youth civic engagement relate to perceptions of school climate? *Applied Developmental Science, 17*(3), 135-147. doi: 10.1080/10888691.2013.804377

Gignac, G. E. (2016). The higher-order model imposes a proportionality constraint: That is why the bifactor model tends to fit better. *Intelligence, 55,* 57-68. doi:10.1016/j.intell.2016.01.006

Graham, J. W. (2012). *Missing data: Analysis and design*. Springer Science & Business Media.

Haste, H., & Hogan, A. (2006). Beyond conventional civic participation, beyond the moral-political divide: Young people and contemporary debates about citizenship. *Journal of Moral Education*, *35*, 473-493. doi:10.1080/03057240601012238

Heider, F. (1958). *The psychology of interpersonal relations.* New York: John Wiley.

Horn, J., & Mcardle, J. J. (1992) A practical and theoretical guide to measurement invariance in aging research, *Experimental Aging Research, 18*:3, 117-144, doi: 10.1080/03610739208253916

Howard, W. J., Rhemtulla, M., & Little, T. D. (2015). Using principal components as auxiliary variables in missing data estimation. *Multivariate Behavioral Research*, (ahead-ofprint), 1-15. doi: 10.1080/00273171.2014.999267

Kahne, J., Middaugh, E., & Schutjer-Mance, K. (2005). California Civic Index [Monograph]. New York, NY: Carnegie Corporation and Annenberg Foundation.

Kahne, J. E., & Sporte, S. E. (2008). Developing citizens: The impact of civic learning opportunities on students' commitment to civic participation. *American Educational Research Journal*, *45*, 738-766. doi:10.3102/0002831208316951

Kirshner, B. (2015). *Youth activism in an era of education inequality*. NYU Press.

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford Publications.

Lenzi, M., Vieno, A., Perkins, D. D., Santinello, M., Elgar, F. J., Morgan, A., & Mazzardis, S. (2012). Family affluence, school and neighborhood contexts and adolescents' civic engagement: A cross-national study. *American Journal of Community Psychology*, *50*, 197-210. doi:10.1007/s10464-012-9489-7

Lerner, R. M., Wang, J., Champine, R. B., Warren, D. J., & Erickson, K. (2014). Development of civic engagement: Theoretical and methodological issues. *International Journal of Developmental Science*, *8*(3), 69-79. doi:10.3233/DEV-14130

Little, T. D. (2013). *Longitudinal structural equation modeling*. New York, NY: Guilford Press.

Little, T. D., Slegers, D. W., & Card, N. A. (2006). A non-arbitrary method of identifying and scaling latent variables in SEM and MACS models. *Structural Equation Modeling: A Multidisciplinary Journal, 13,* 59-72. doi: 10.1207/s15328007sem1301_3

Mahatmya, D., & Lohman, B. J. (2012). Predictors and pathways to civic involvement in emerging adulthood: Neighborhood, family, and school influences. *Journal of Youth and Adolescence, 41*, 1168-1183. doi:10.1007/s10964-012-9777-4

Malin, H., Ballard, P. J., & Damon, W. (2015). Civic purpose: An integrated construct for understanding civic development in adolescence. *Human Development*, *58*(2), 103-130.

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology*, *93*(3), 568. doi: 10.1037/0021-9010.93.3.568

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525-543. doi: 10.1007/BF02294825

Metzger, A., & Ferris, K. (2013). Adolescents' domain-specific judgments about different forms of civic involvement: Variations by age and gender. *Journal of adolescence*, *36*(3), 529-538. doi: 10.1016/j.adolescence.2013.03.003

Metzger, A., Oosterhoff, B., Palmer, C. A., & Ferris, K. (2014). Dimensions of citizenship: Associations among adolescents' sociopolitical values and civic judgments. *PS: Political Science & Politics*, *47*(02), 443-448.

Metzger, A., & Smetana, J. G. (2009). Adolescent civic and political engagement: Associations between domain-specific judgments and behavior. *Child Development*, *80*, 433-441. doi:10.1111/j.1467-8624.2009.01270.x

National Center for Education Statistics. (2014). http://nces.ed.gov/ccd/schoolsearch/ Retrieved April 30, 2014.

Overton, W. F. (2015). Processes, relations, and relational-developmental-systems. In R. M. Lerner (Ed.), W. F. Overton & P. C. M. Molenaar (Vol. Eds.), *Handbook of child psychology and developmental science* (7th edition, pp. 9-62). Hoboken, NJ: Wiley. doi: 10.1002/9781118963418.childpsy102

Padilla-Walker, L. M., & Carlo, G. (Eds.). (2014). *Prosocial development: A multidimensional approach*. New York, NY: Oxford University Press.

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47,* 667–696. doi: 10.1080/00273171.2012.715555

Rothbart, M. K. (2012). Advances in temperament. In M. Zentner & R. L. Shiner (Eds.), *Handbook of temperament* (pp. 3-20). New York, NY: The Guilford Press.

Ryff, C. D. (2014). Psychological well-being revisited: Advances in science and practice. *Psychotherapy and Psychosomatics, 83*(1), 10-28. doi:10.1159/000353263

Satorra, A. & Bentler, P. (1999, August). *A scaled difference chi-square test statistic for moment structure analysis.* UCLA Statistics working paper #260. Retrieved on April 4, 2016 from http://preprints.stat.ucla.edu/download.php?paper=260

Sherrod, L. R., & Lauckhardt, J. (2009). The development of citizenship. In R. M. Lerner & L. Steinberg (Eds.), *Handbook of adolescent psychology: Contextual influences on adolescent development, Vol. 2* (3rd ed., pp. 372–408). Hoboken, NJ: Wiley.

Syvertsen, A. K., Wray-Lake, L., & Metzger, A. (2015). *Youth civic and character measures toolkit.* Minneapolis, MN: Search Institute.

Torney-Purta, J., Cabrera, J. C., Roohr, K. C., Liu, O. L., & Rios, J. A. (2015). Assessing civic competency and engagement in higher education: Research background, frameworks, and directions for next-generation assessment. *ETS Research Report Series*, *2015*(2), 1-48. doi: 10.1002/ets2.12081

Voight, A., & Torney-Purta, J. (2013). A typology of youth civic engagement in urban middle schools. *Applied Developmental Science, 17*, 198-212. doi:10.1080/10888691.2013.836041

Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child*

Development Perspectives, 4(1), 10-18. doi: 10.1111/j.1750-8606.2009.00110.x

Wray-Lake, L., Rote, W., Victorino, C., & Benavides, C.M. (2014). Examining developmental transitions in civic engagement across adolescence: Evidence from a national U.S. sample. *International Journal of Developmental Science, 8,* 95-104. doi: 10.3233/DEV-14142

Wray-Lake, L., & Sloper, M. (2015). Youth civic engagement and the role of experiences across contexts. *Applied Developmental Science.* doi: 10.1080/10888691.2015.1114888

Wray-Lake, L., Syvertsen, A.K., & Flanagan, C.A. (2016). Developmental change in social responsibility during adolescence: An ecological perspective. *Developmental Psychology, 52(1)*, 130-142. doi: 10.1037/dev0000067

Zaff, J. F., Boyd, M., Li, Y., Lerner, J. V., & Lerner, R. M. (2010). Active and engaged citizenship: Multi-group and longitudinal factorial analysis of an integrated construct of civic engagement. *Journal of Youth and Adolescence, 39*, 736-750. doi: 10.1007/s10964-010-9541-6

Table 1. *Conceptual model comparison.*

Approach	Purpose	Theoretical Value	Practical Value	Major Disadvantage
Model 0: Unidimensional	Views civic engagement as unidimensional and offers only shared variance to be predicted	Promotes holistic understanding and broad generalizable knowledge of a construct	Leads to simpler and more straightforward "global" recommendations	Ignores all variations in specific dimensions, viewing them as error in the model
Model 1: Correlated Unidimensional Factors	Emphasizes and predicts differences among dimensions of a construct	Allows for testing and drawing conclusions about specificity in processes	Provides precision in recommendations about specific pathways to specific constructs	High degree of model complexity and thus higher probability of Type-I error; potential for post-hoc rather than a priori explanations
Model 2: Higher-Order Factor	Predicts and understands shared variance among dimensions of a construct	Promotes holistic understanding and broad generalizable knowledge of a construct	Leads to simpler and more straightforward "global" recommendations	Masks potentially important variations in specific dimensions
Model 3: Bifactor	Predicts differences and models shared variance among dimensions	Acknowledges specific and general processes leading to integrated understanding of a construct	Offers a balance of general and specific recommendations for practice	High degree of model complexity; must use caution in interpreting general and specific pieces of model

Table 2. *Correlations among civic engagement items.*

	Social Responsibility			Informal Helping						Political Beliefs		Civic Skills						Enviro. Behavior			Manifest Variables		
	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.	21.	22.	23.	24.
Social Resp. Values																							
1. …others' needs	.57	.47	.38	.27	.29	.31	.18	.26	.15	.29	.22	.31	.30	.28	.27	.30	.29	.16	.17	.13	.23	.27	.14
2. …help needy	---	.50	.43	.29	.26	.32	.19	.26	.21	.25	.22	.25	.30	.24	.25	.24	.27	.13	.16	.14	.22	.22	.10
3. …fair treatment		---	.44	.22	.22	.28	.14	.18	.16	.26	.18	.22	.29	.26	.23	.29	.33	.19	.18	.14	.13	.22	.10
4. …actions			---	.18	.21	.22	.14	.15	.14	.23	.19	.21	.27	.24	.24	.25	.29	.19	.20	.16	.11	.18	.13
Informal Helping																							
5. …stood up				---	.35	.34	.18	.37	.28	.16	.13	.25	.30	.27	.27	.22	.23	.08	.09	.08	.27	.13	.13
6. … homework					---	.47	.26	.32	.24	.22	.16	.35	.34	.30	.31	.30	.30	.04	.12	.03	.24	.25	.13
7. …supplies						---	.28	.29	.23	.25	.18	.24	.28	.24	.26	.26	.29	.11	.15	.08	.18	.18	.12
8. …help at home							---	.32	.24	.11	.08	.17	.17	.17	.14	.16	.16	.05	.09	.09	.13	.13	.08
9. …neighbors								---	.47	.19	.16	.21	.25	.22	.27	.20	.19	.10	.12	.13	.32	.17	.16
10. …babysit									---	.12	.12	.15	.16	.18	.22	.14	.13	.02	.10	.09	.25	.08	.09
Political Beliefs																							
11. … news										---	.45	.27	.27	.28	.28	.30	.29	.10	.16	.11	.17	.35	.25
12. …protests											---	.20	.25	.20	.22	.20	.19	.08	.15	.08	.13	.20	.16
Civic Skills																							
13. …create plan												---	.54	.56	.49	.53	.52	.11	.12	.08	.23	.40	.16
14. …others care													---	.56	.53	.54	.50	.15	.15	.11	.23	.29	.13
15. …say views														---	.54	.57	.55	.11	.10	.12	.17	.35	.14
16. …leaders															---	.56	.52	.10	.10	.15	.23	.28	.22
17. …hear views																---	.60	.09	.09	.08	.20	.35	.18
18. …summarize																	---	.15	.12	.10	.16	.34	.17
Enviro. Behavior																							
19. ...electronics																		---	.43	.41	.03	.08	.07
20. …paper use																			---	.50	.07	.09	.12
21. …water use																				---	.10	.01	.10
Manifest Variables																							
22. Volunteering																					---	.17	.18
23. Voting Intentions																						---	.23
24. News Consumption																							---

Note. All bivariate correlations were significant at $p \leq .05$, with the exception of those in shaded cells.

Table 3. *Final Model Fit Indices and Model Comparisons.*

MODEL FIT INDICES							
MODEL	**MLR χ^2**	**CFI**	**TLI**	**RMSEA**	**SRMR**	**AIC**	**BIC**
Model 0: Unidimensional	3659.01(252) , $p < .001$	.692	.662	.074 (.072-.076)	.080	153748.93	154318.38
Model 1: Correlated Unidimensional	831.29(228) , $p < .001$	.945	.934	.033 (.030 -.035)	.033	150445.09	151154.01
Model 2: Higher-Order	1012.24(248), $p < .001$	.931	.923	.035 (.033-.038)	.039	150618.59	151211.29
Model 3: Bifactor	870.35(232) , $p < .001$	.942	.931	.033 (.031-.036)	.036	138163.44	138698.03

MODEL COMPARISONS		
	S-B $\Delta\chi^2$(df)	**ΔCFI**
Unidimensional vs.		
Correlated Unidimensional	2520.41(24)***	.253
Higher-Order	1579.65(4)***	.014
Bifactor	2394.46(20)***	.250
Correlated Unidimensional vs.		
Higher-Order	-178.18(20)***	.014
Bifactor	-41.15(4)***	.003
Higher-Order vs.		
Bifactor	137.74(24)***	.011

Note. S-B = Satorra-Bentler Scaled X^2 difference test. CFI = Comparative fit index. TLI = Tucker-Lewis index. RMSEA = Root mean square error of approximation. SRMR= Standardized root mean square residual.

Table 4. *Factor loadings for all measurement structural equation models.*

	Model 0: Unidimensional		**Model 1: Correlated Unidimensional**		**Model 2: Higher-Order**				**Model 3: Bifactor**			
					1st order		2nd order		1st order		General	
	B	SE	B	SE	B	SE	B	SE	B	SE	B	SE
Social Resp. Values							.722	.018				
1. …others' needs	.579	.018	.727	.017	.730	.017			.461	.028	.553	.020
2. …help needy	.568	.027	.758	.028	.757	.028			.573	.029	.517	.029
3. …fair treatment	.522	.020	.662	.024	.662	.024			.481	.032	.457	.025
4. …actions	.474	.022	.575	.027	.574	.027			.397	.043	.413	.024
Informal Helping							.749	.030				
5. …stood up	.461	.031	.574	.026	.565	.027			.282	.051	.465	.032
6. …homework	.518	.035	.616	.032	.625	.032			.175	.058	.565	.042
7. …help at home	.507	.034	.598	.030	.609	.029			.160	.072	.549	.040
8. …supplies	.329	.029	.427	.028	.432	.028			.290	.045	.317	.036
9. …neighbors	.449	.033	.617	.028	.608	.028			.600	.032	.426	.033
10. …babysit	.347	.032	.499	.031	.489	.029			.542	.032	.296	.034
Political Beliefs							.672	.031				
11. …news	.466	.025	.722	.018	.717	.017			.545	.024	.492	.030
12. …protests	.366	.020	.632	.017	.633	.017			.495	.022	365	.020
Civic Skills							.760	.023				
13. …create plan	.687	.025	.719	.028	.716	.028			.438	.048	.565	.023
14. …others care	.711	.026	.723	.021	.725	.021			.434	.024	.579	.024
15. …say views	.708	.025	.755	.024	.753	.024			.538	.037	.536	.025
16. …leaders	.695	.026	.716	.026	.719	.024			.459	.032	.028	.028
17. … hear views	.717	.026	.763	.015	.763	.015			.541	.044	.547	.044
18. …summarize	.703	.035	.732	.032	.731	.032			.487	.028	.548	.038
Enviro. Behavior							340	.037				
19. ...electronics	.226	.026	.593	.026	.592	.025			.557	.029	.202	.028
20. …paper use	.226	.025	.734	.040	.734	.043			.674	.050	.259	.022
21. …water use	.277		.686	.033	.686	.032			.678	.036	.181	.040
22. **Volunteering**	.365	.023	---	---	---	---	.410	.022			.398	.023
23. **Voting**	.453	.024	---	---	---	---	.496	.024			.483	.024
24. **News**	.266	.033	---	---	---	---	.328	.039			.317	.038

Note. All factor loadings were significant at $p \leq .001$.

Table 5. *Correlations and means from the correlated unidimensional factors model.*

	Means	Correlations						
		2.	3.	4.	5.	6.	7.	8.
1. Social Responsibility Values	3.95	.587	.498	.522	.329	.266	.326	.165
2. Informal Helping	3.33	---	.429	.574	.237	.416	.289	.215
3. Political Beliefs	3.32		---	.498	.264	.224	.425	.310
4. Civic Skills	3.69			---	.224	.277	.454	.213
5. Environmental Behavior	3.06				---	.099	.085	.150
6. Volunteering	2.67					---	.173	.172
7. Voting Intentions	3.65						---	.229
8. News Consumption	3.15							---

Note. All bivariate correlations were significant at $p \leq .05$. Latent means were derived using effects coding.

Table 6. *Models linking civic engagement models to purpose.*

	PURPOSE		
	Model 1: Correlated Unidimensional	**Model 2: Higher-Order**	**Model 3: Bifactor**
Social Resp. Values	.235(.06)***	---	.169(.05)***
Informal Helping	.266(.04)***	---	.150 (.05)**
Political Beliefs	.111(.04)**	---	.072(.04)
Civic Skills	.243(.07)***	---	.158(.05)***
Enviro. Behavior	-.055(.06)	---	-.045(.06)
Volunteering	.011(.02)	---	.016(.03)
Voting Intentions	.063(.03)*	---	.075(.04)*
News Consumption	.033(.03)	---	.023(.03)
Civic Engagement	---	.775(.03)***	.608(.07)***
Model Fit	MLR $\chi^2(318) = 980.51$, $p < .001$, CFI = .95, TLI = .93, RMSEA = .031 (.029 - .033), SRMR = .034	MLR $\chi^2(42) = 1175.91$, $p < .001$, CFI = .93, TLI .93, RMSEA = .033 (.031 - .035), SRMR = .040	MLR $\chi^2(295) = 1026.47$, $p < .001$, CFI = .94, TLI = .93, RMSEA = .032 (.030 - .034), SRMR = .036

Note. Standardized coefficients are reported. Standard errors in parentheses. *$p \leq .05$, **$p \leq .01$, **$p \leq .001$. 90% confidence intervals shown in parentheses for RMSEA.

Table 7. *Measurement invariance tests and latent mean differences by age.*

MEASUREMENT INVARIANCE TESTS					
	MLR χ^2(df)[a]	**CFI**	**RMSEA**	**Δ CFI**	**S-B Δχ^2(df)[b]**
Model 1: Correlated Unidimensional					
Configural (Baseline) Model	1707.38(685)	.933	.043		
Metric Invariance	1801.62(716)	.928	.043	.005	100.61(31)***
Scalar Invariance	1916.96(748)	.923	.044	.005	95.28(32)***
Model 2: Higher-Order					
Configural (Baseline) Model	1977.50(744)	.919	.045		
Metric Invariance – Lower Order	2082.52(774)	.914	.045	.005	118.08(30)***
Metric Invariance – Higher Order	2092.60(778)	.913	.045	.001	9.94(4)*
Scalar Invariance – Lower Order	2179.55(816)	.910	.045	.003	96.40(38)***
Scalar Invariance – Higher Order	2457.63(824)	.892	.049	.018	1027.32(8)***
Partial Scalar Invariance – Higher Order[c]	2144.43(820)	.913	.044	.003	2.21(4)
First Order Disturbances	2144.72(829)	.913	.044	.000	1.48(9)
Model 3: Bifactor					
Configural (Baseline) Model	1686.72(697)	.935	.042		
Metric Invariance	1865.38(772)	.928	.042	.007	178.67(75)***
Scalar Invariance	1994.29(808)	.922	.042	.006	108.05(36)***
First Order Disturbances	2013.84(820)	.921	.042	.001	21.72(12)*

[a] All model X2s were significant at $p < .001$.[b] Satorra-Bentler Scaled X^2 difference test. [c] Partial invariance achieved (first order mean differences shown in Table 8).

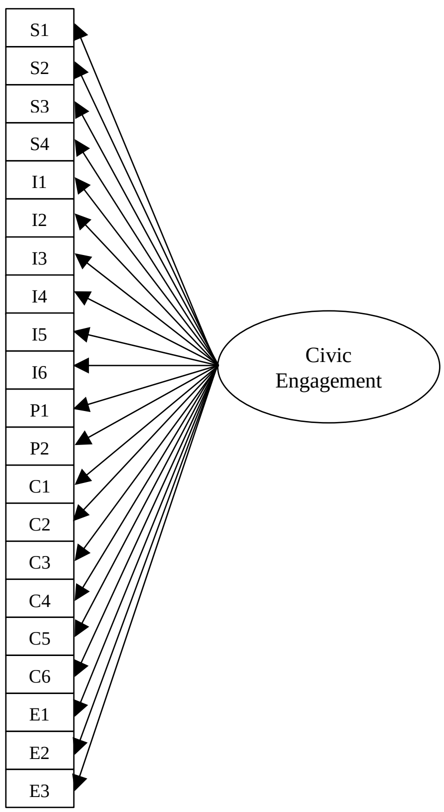
Table 8. *Age differences in levels of civic engagement*

	Model 1: Correlated Unidimensional			**Model 2: Higher-Order**			**Model 3: Bifactor**		
	Elementary	Middle	High	Elementary	Middle	High	Elementary	Middle	High
Social Responsibility Values	.00	.00	.00	.00	.00	-.25	.00	-.27	-.54
Informal Helping	.00	.37	.37	.00	.44	.44	.00	.00	.00
Political Beliefs	.00	.00	.20	.00	.00	.00	.00	.00	.00
Civic Skills	.00	.00	.32	.00	.00	.24	.00	.00	.00
Environmental Behavior	.00	-.58	-1.13	.00	-.66	-1.34	.00	-.71	-1.45
Volunteering	2.66	2.66	2.66	2.55	2.55	2.55	2.53	2.53	2.53
Voting Intentions	3.66	3.66	3.66	3.55	3.55	3.55	3.52	3.52	3.52
News Consumption	3.15	3.15	3.15	3.07	3.07	3.07	3.06	3.06	3.06
Civic Engagement	--	--	--	.00	.00	.39	.00	.00	.47

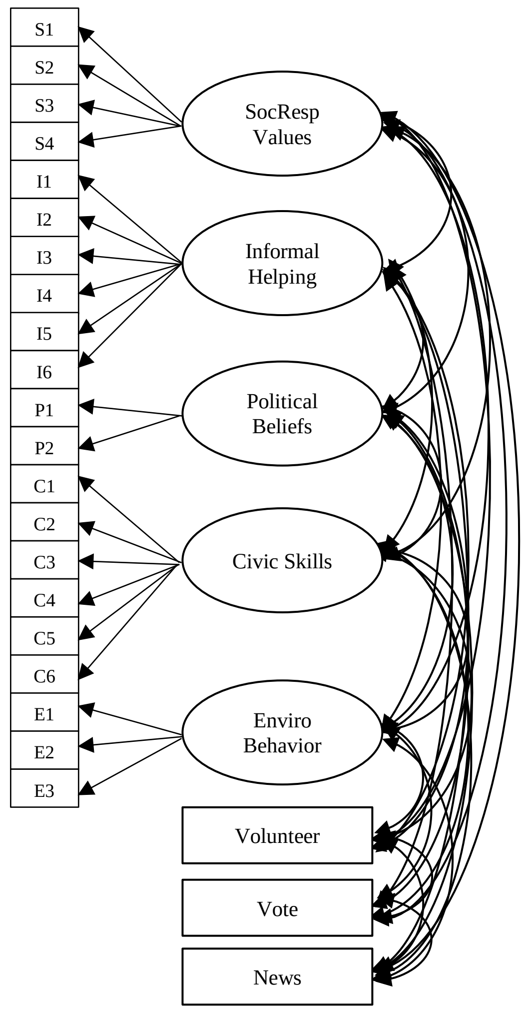
Note. Latent factor means are estimated as relative differences from elementary-aged youth, fixed at 0. Middle and high school youths' latent means were compared by constraining means to be equal and chi-square tests assessed differences. Values in rows within each model that differ are significant at p <.002 and values that are the same were not significantly different and constrained to be equal.

Figures 1-4. *Conceptual figures illustrating various multidimensional measurement model approaches.*

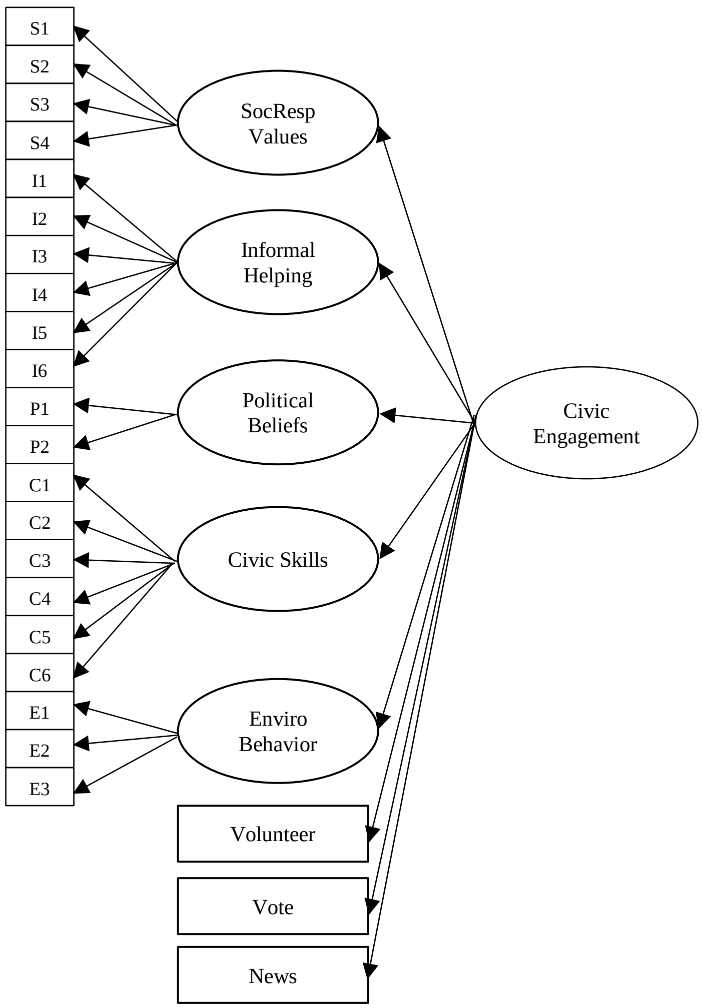
Model 0. Unidimensional Model



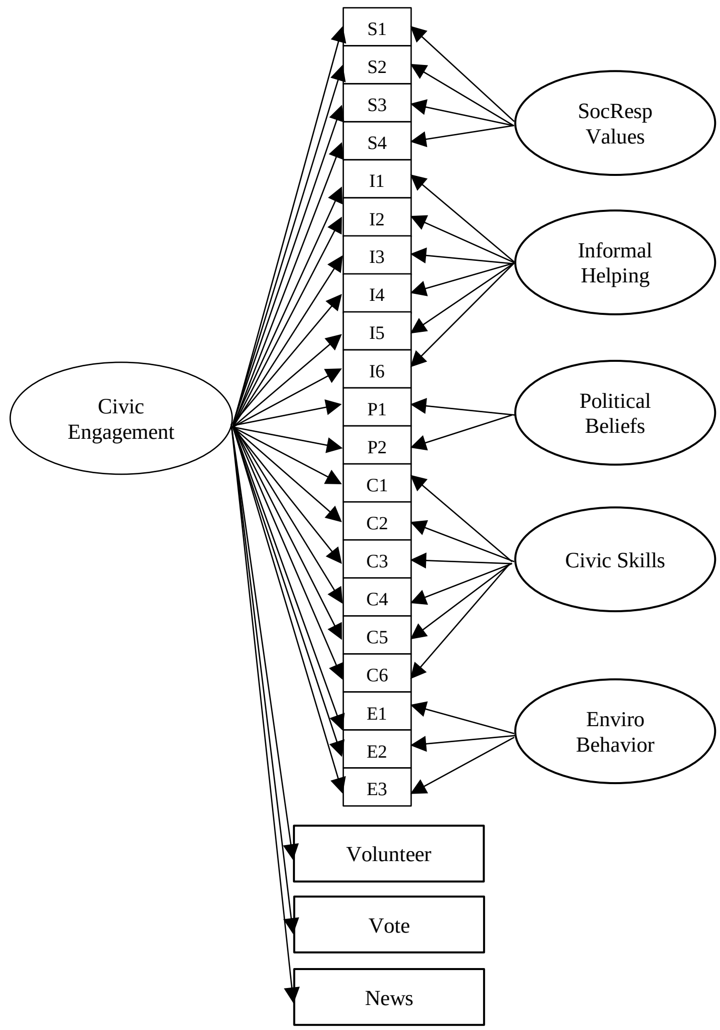
Model 1. Correlated Unidimensional Factors Model



Model 2. Higher-Order Model



Model 3. Bifactor Model



Online Appendix A

Description of Planned Missingness Design

Data collection for the Roots of Engaged Citizenship Project used a planned missingness design. Planned missingness designs represent an efficient way to maximize the number of survey questions asked in a fixed time frame. This design reduces cognitive demands on participants, produces surveys that are developmentally responsive to participants' abilities, and increases likelihood of survey completion, thus minimizing less desirable forms of missing data (Enders, 2010; Graham, 2012; Little & Rhemtulla, 2013). Our study assessed multiple individual and contextual factors in relation to various dimensions of civic engagement. To achieve our aims, we used a school-based survey design where students were surveyed for approximately 45 minutes during school hours. The planned missingness design allowed us to include a wider set of constructs while keeping the survey length short enough to complete during a class period.

A three-form planned missing design was employed, such that survey items were divided into one core set (X) and three additional item sets (A, B, C). Three survey forms were created that included the core items (X) and two out of three other items sets (see Table A1). The X set was presented first in all survey versions, and item sets A, B, and C were counterbalanced across versions. Items in a scale were kept together in the same item set (Graham, 2012). The core X set contained the primary dependent variables for the study (i.e., civic engagement), demographics, and a few central predictors. A, B, and C sets were structured to include a balance of competencies, character, and context variables across item sets. Consistent with recommendations for planned missingness designs, constructs were grouped together within version if we had specific hypotheses about how they were associated, which maximized power to test these hypotheses (Graham, 2012).

Table A1. *Survey forms and item sets*

Survey Form	**Item Sets**	**Number of Items**		
		Elementary	**Middle**	**High**
1	X + A + B	96	153	173
2	X + C + A	96	149	169
3	X + B + C	96	150	167
	Total # of Items Measured	134	225	247

Another layer of our design was planned missingness by age. After creating the three forms, we created age-specific versions of the survey such that elementary and middle school versions were shorter than the high school version (see Table A1). Item wording was the same across ages, but more complex and less central constructs were included at older ages only. This resulted in 9 versions of the survey (3 forms X 3 age groups). The civic engagement measurement models utilized items that were measured identically across all three age groups.

Given that this type of missing data is completely controlled by the researcher and thus missing completely at random (MCAR), modern missing data approaches can easily accommodate this form of missingness. The Principal Components Method was used to handle planned and other types of missing data (Howard, Rhemtulla, & Little, 2015; Little, Howard, McConnell, & Stump, 2008). PCA is conducted on all variables in the data and resulting principal components are used as auxiliary variables in conjunction with full information maximum likelihood (FIML) estimation. Auxiliary variables improve the performance of FIML by making assumptions of missing at random (MAR) more reasonable and increasing FIML efficiency by reducing uncertainty due to missingness (Collins, Schafer, & Kam, 2001). PCA was conducted using the Quark package for R version 3.1.2. For each school level (elementary, middle, and high), variables were first standardized in Quark to ensure that all variables contributed equally to PC scores. Following standardization, a single imputation was run so that all variables were included in the PCA. After conducting the PCA on the imputed data, 10 of the initial PCs were retained, accounting for 50% of the variance. Sensitivity analyses determined the number of principal components to ensure that the number of retained PCs provided consistent results across models. These PCs were merged with the original, non-imputed data and used as auxiliary variables in FIML estimations in M*plus* version 7.

A series of chi-square tests were conducted to determine whether survey versions were distributed randomly across participants. We examined survey version in relation to site (California, Minnesota, West Virginia), school level (elementary, middle, and high), grade (4^{th}-12^{th}), gender (male, female), ethnicity (Black, Hispanic, White, Asian, Other), and immigrant status (born in US, born outside of the US). As shown in Table A2, no chi-square tests were significant, indicating that assignment to survey version did not vary by demographics. This provides evidence that the planned missing design was successfully executed and randomly distributed across participants.

Table A2. *Chi-square tests of independence with survey version and key demographics.*

Pearson chi-square tests	χ^2	**df**	***p*-value**
Survey version (3) x Site (3)	1.60	4	.809
Survey version (3) x School level (3)	.947	4	.918
Survey version (3) x Grade (9)	2.50	16	1.00
Survey version (3) x Gender (2)	2.06	2	.357
Survey version (3) x Ethnicity (5)	11.08	12	.523
Survey version (3) x Immigrant status (2)	1.16	2	.559

Appendix References

Collins, L. M., Schafer, J. L., & Kam, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, *6*(4), 330. doi: 10.1037//1082-989X.6.4.330

Enders, C. K. (2010). *Applied missing data analysis*. New York: Guilford Press.

Graham, J. W. (2012). *Missing data: Analysis and design*. New York: Springer Science & Business Media.

Howard, W., Rhemtulla, M., & Little, T. D. (2015). Using principal components as auxiliary variables in missing data estimation. *Multivariate Behavioral Research*, *50*(3), 285-299. doi: 10.1080/00273171.2014.999267.

Little, T. D., McConnell, E. K., Howard, W. J., & Stump, K. N. (2008). *Missing Data in Large Data Projects: Two methods of missing data imputation when working with large data projects.* (KUant Guides#11.3). Retrieved from the Center for Research and Data Analysis Website: http://web.ku.edu/~quant/cgi-bin/mw0/index.php?title=KUant_Guides

Little, T. D., & Rhemtulla, M. (2013). Planned missing data designs for developmental researchers. *Child Development Perspectives*, *7*(4), 199-204. doi: 10.1111/cdep.12043